

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 15, 2003, 03:45:56 ; Search time 2149 Seconds
(without alignments)
3751.269 Million cell updates/sec

Title: US-09-600-932-2

Perfect score: 1484
Sequence: 1 MNGFASLRRNQFILLVFL.....NDTECHLNYFVCEFIKKKK 277

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+P2n.model -DEV=xlp
-Q/cgn2.1/USPTO.spool/US09600932.r/unat_09062003_094530_25981/app.query.fasta_1.455
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -IOOPCL=0 -IOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdt -LIST=45
-DOCLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-USER=US09600932.ecn_1.1652-@runat_09062003_094530_25981 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEDEV -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mem:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1484	100.0	1594	AB002631	AB002631 Homo sapi
2	1484	100.0	1595	E29008	E29008 Novel colle
3	1472	99.2	1016	AX376032	AX376032 Sequence
4	720.5	48.6	141262	AC080033	AC080033 Homo sapi
5	720.5	48.6	182475	AC023487	AC023487 Homo sapi
6	719.5	48.5	1238	AX403469	AX403469 Sequence
7	719.5	48.5	1238	AX454582	AX454582 Sequence
8	719.5	48.5	1238	AX491060	AX491060 Sequence
9	719.5	48.5	1257	BC000078	BC000078 Homo sapi
10	682	46.0	267118	AC097055	AC097055 Rattus no
11	673	45.4	167718	AC115924	AC115924 Mus muscu
12	459	30.9	139357	AC010907	AC010907 Homo sapi
13	351	23.7	1310	AY071821	AY071821 Bos tauri
14	349.5	23.6	1454	BTLSPD	BTLSPD
15	344.5	23.2	1183	BC003705	BC003705 Mus muscu
16	344.5	23.2	1253	MUSSPD	MUSSPD
17	344	23.2	1095	BTCA3	BTCA3
18	338.5	22.8	1385	AF132496	AF132496 Sus scrof
19	337	22.7	1410	AX334792	AX334792 Sequence
20	337	22.7	1410	HSNRPD	HSNRPD
21	337	22.7	1559	AF231714	AF231714 Gallus ga
22	334.5	22.5	1000	AF227738	AF227738 Danio rer
23	333	22.4	1265	RARSPP	RARSPP
24	333	22.4	1301	BC022318	BC022318 Homo sapi
25	332.5	22.4	1305	BTCONGL	BTCONGL
26	330.5	22.3	1284	AF509590	AF509590 Bos tauri
27	328.5	22.1	1385	BOVCGT	BOVCGT
28	328.5	22.1	1392	E16477	E16477 Artificial
29	328.5	22.1	1519	BOVCGSLUTA	BOVCGSLUTA
30	327.5	22.1	714	AF022226	AF022226 Gallus ga
31	327.5	22.1	714	AR182170	AR182170 Sequence
32	316	21.3	1054	AF227737	AF227737 Cyprinus
33	307	20.7	934	AF227739	AF227739 Carassius
34	305.5	20.6	2967	RABSPARO	RABSPARO
35	298	20.1	1052	AF211856	AF211856 Ovis arie
36	298	20.1	1901	AF076633	AF076633 Ovis arie
37	297.5	20.0	805	MACMBPC	MACMBPC
38	294	19.8	888	AB015963	AB015963 Equus cab
39	293.5	19.8	1010	E37364	E37364 Swine serum
40	291	19.6	799	MACMBPA	MACMBPA
41	290	19.5	1069	MUSRRPA	MUSRRPA
42	290	19.5	1098	BC010760	BC010760 Mus muscu
43	289.5	19.5	747	E27637	E27637 Recombinant
44	289.5	19.5	776	AF360991	AF360991 Homo sapi
45	289.5	19.5	900	AR182149	AR182149 Sequence

RESULT 1

ALIGNMENTS

AB002631 1594 bp mRNA linear PRI 24-JUN-1999
 LOCUS Homo sapiens mRNA for collectin 34, complete cds.
 DEFINITION
 ACCESSION AB002631.1 GI:5162874
 VERSION collectin 34.
 KEYWORDS Homo sapiens cDNA to mRNA.
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Chiroptera; Carnivora; Homnidae; Homo.
 1 (sites)
 Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Keshi, H., Sakai, Y., Fukuchi, A., Sakamoto, T., and Wakamaya, N., Molecular cloning of a novel human collectin from liver (CL-L1) J. Biol. Chem. 274 (19), 13681-13689 (1999)
 2 (bases 1 to 1594)
 Ohtani, K.
 Direct Submission
 Submitted (04-APR-1997) Katsuki Ohtani, Osaka Prefectural Institute of Public Health, Department of Pathology, 3-69, Nakamichi 1-chome Higashinari-ku, Osaka, Osaka 537, Japan
 (E-mail: suzuki@iph.pref.osaka.jp, Tel: +81-6-972-1321, Fax: +81-6-972-0772)
 FEATURES
 source
 location/Qualifiers
 1..1594
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 6..839
 /codon_start=1
 /product="collectin 34"
 /protein_id="BA81747.1"
 /db_xref="GI:5162875"
 /translation="MNGFASLRLRNOFILVFLQIQSLDIDSPRAEVCATHTI SPSPKGDGEGDPGEGRKGVGRMGPKIKGELGDMGRNGIKTGPICKGDKGE KGLGIPGEKAGTVCDCGRYRFVQGLDSTARLTKSMKPFYNAVIGARETEREY YVQEKYRESLTHCRIRGMLAMPDEANLIIAYVAKSGFERFVIGVNDIEREG QYVFTDPTPLQNSNMNEGEPSDPYGHEDCEVHLSRRMDTECHLMYFCETIKK K."
 polyA_signal 1037..1041
 BASE COUNT 444 a 321 c 382 g 447 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,78e-115 Length: 1594
 Score: 1484.00 Matches: 277
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0
 US-09-600-932-2 (1-277) x AB002631 (1-1594)
 QY 1 MetAsnGlyPheAlaSerLeuLeuArgArgAsnGlnPheIleLeuLeuValLeuPheLeu 20
 DB 6 ATGAATGGCTTTGGATCCCTTCGTCGAAAGAACCAATTATCCCTCGTACATTTCTT 65
 QY 21 LeuGlnIleGlnSerLeuGlyLeuAspIleAspSerArgProThrAlaGluValCysAla 40
 DB 66 TTGCAAAATTCAGAGTGGGCTGATATGATAGCCGCTCCTACCGCTGAAGTGTGGCC 125
 QY 41 ThrHisIleSerProGlyProGlyAspAspGlyGlyGlyGlyGlyGlyGlyGlyGly 60
 DB 126 ACACACACAAATTTACACAGACCCCAAGAGATGATGTAAGAAAGAGATCCAGAGAA 185
 QY 61 GluGlyLysHisGlyLysValGlyArgMetGlyProLysGlyIleLysGlyGluLeuGly 80
 DB 186 GAGGAGAAAGCATGGCAAGTGGGACGATGGGCGCAAGAAAGAAATTAAGAGAACTGGGT 245
 QY 81 AspMetGlyAspArgGlyAsnIleGlyLysThrGlyProIleGlyLysGlyAspLys 100
 DB 246 GATATGGAGATCGGGCAATATTGGCAAGACAGCTGGCCCATTTGGCAAGAGGTGACAAA 305

QY 101 GlyLysGlyLeuLeuGlyIleProGlyGlyLysGlyLysAlaGlyThrValCysAsp 120
 DB 306 GGGGAAAAAGGTTGCTTGGATACCTGGAGAAAAAGCAAGCAAGCAAGTGTGTGAT 365
 QY 121 CysGlyArgThrArgLysPheValGlyGlnLeuAspIleSerIleAlaArgLeuLysThr 140
 DB 366 TGTGAAGATACCGGAAATTTGTTGGACAACTGATATTGATTTGCCGCCGCTCAAGACA 425
 QY 141 SerMetLysPheValLysAsnValIleAlaGlyIleArgGlnThrGlnGlyLysPheThr 160
 DB 426 TCTATGAAGTTTGCAGAAATGTATGACAGGATTTGGGAAACTGAAGAAATTTTAC 485
 QY 161 TyrIleValGlnGluGlyLysAsnThrArgGluSerLeuThrHisCysArgIleArgGly 180
 DB 486 TACATCGTGCAGGAAGGAAGAACTACAGGAAATCCCTAACCCCTCAGGATTCGGGGT 545
 QY 181 GlyMetLeuAlaMetProLysAspGluAlaAlaAsnThrLeuIleAlaAspTyrValAla 200
 DB 546 GGAATGCTAGCCATGCCAAGAGATGAAGCTGCCAACACACTCATCGCTGACTATGTGCC 605
 QY 201 LysSerGlyPhePheArgValPheIleGlyValAsnAspLeuGluArgGlyGlnTyr 220
 DB 606 AAGATGGCTTCTTTCGGGTGTTTATGCGGTGAATGACCTTGAAAGGAGGACATAC 665
 QY 221 MetPheThrAspAsnThrProLeuGlnAsnTyrSerAsnTyrAsnGluGlyProSer 240
 DB 666 ATGTTACAGACAAACACTCCACTCAGAACTATAGCAACTGAAATGAGGGGAAACCCAGC 725
 QY 241 AspProTyrGlyHisGluAspCysValGluMetLeuSerSerGlyArgTyrPasnAspThr 260
 DB 726 GACCCCTATGCTCATGAGAGACTGTGTGAGATGCTGTGACCTGTGCAATGAAATGACACA 785
 QY 261 GluCysHisLeuThrMetTyrPheValCysGluPheIleLysLysLysLys 277
 DB 786 GAGGCGCATCTTACCATGATCTTGTCTGAGATTCATCAAGAAAGAAAG 836
 RESULT 2
 E29008
 LOCUS Novel collectin. 1595 bp DNA linear PAT 18-JUN-2001
 DEFINITION
 ACCESSION E29008
 VERSION E29008.1 GI:13018416
 KEYWORDS JP 1999206377-A/1.
 SOURCE
 ORGANISM
 Homo sapiens.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.
 1 (bases 1 to 1595)
 REFERENCE
 AUTHORS Nobutaka, W.
 TITLE Novel collectin
 JOURNAL Patent: JP 1999206377-A 1 03-AUG-1999;
 COMMENT
 FUSO YAKOHIN KOGYO KK
 OS Homo sapiens (human)
 PN JP 1999206377-A/1
 PD 03-AUG-1999
 PF 23-JAN-1998 JP 1998011281
 PR NOBUTAKA WAKAMAYA
 PC C12N15/09, C07K14/47, C07K14/78, C12P21/00, C12N15/00 CC
 Strandedness: Double;
 CC Topology: Linear;
 FH key
 FT CDS location/Qualifiers
 1..1595
 location/Qualifiers
 6..836.
 FEATURES
 source
 location/Qualifiers
 1..1595
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 444 a 322 c 382 g 447 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,78e-115 Length: 1595
 Score: 1484.00 Matches: 277

Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%

Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-600-932-2 (1-277) x E29008 (1-1595)

1 MetanaglYPhealSerLeuLeuArgArgAnGlnPheleLeuLeuValLeuPheLeu 20
6 ATGAAATGGCTTTCATCTCTGCTGGAAGAAACCAATTATCTCTCTGTAATTTCTT 65
21 LeuGlnIleGlnSerLeuGlyLeuAspIleAspSerArgProThrAlaGluValCysAla 40
66 TTGCAAAATTCAGAGCTGGCTGCTGATATGATACCCCTCCACCGCTGAAGTCTGTGCC 125
41 ThrHisThrIleSerProGlyProGlyAspAspGlyGlyGlyGlyGlyGlyGlyGly 60
126 ACACACACAAATTCACCAAGACCCAAAGAGATGATGATGATGATGATGATGATGATGAT 185
61 GluGlyLysHisGlyLysValGlyArgMetGlyProGlyGlyLysGlyGlyGlyGlyGly 80
186 GAGGAAAGACATGCGAAAGTGGAGCATGGGGCCGAAAGAAATTAAGAGAGAACTGGGT 245
81 AspmetGlyAspArgGlyAsnIleGlyLysThrGlyProIleGlyLysGlyLysAspLys 100
246 GATATGGGAGATCGGGCAATATGCAAGACTGGGCCCATTTGGGAGAGAGAGGTGACAAA 305
101 GlyLysLysGlyLeuLeuGlyLysProGlyLysGlyLysGlyLysGlyLysGlyLys 120
306 GGGGAAAGAGTTTCTTGGATACCTGAGAAAGAAAGCAAGCAAGTACTCTGTGTGAT 365
121 CysGlyArgTyrArgLysPheValGlyGlnLeuAspIleSerIleAlaArgLeuLysThr 140
366 TGTGAAGATACCGGAAATTTGTTGACACATGATATGATATGATATGATATGATATGAT 425
141 SerMetLysPheValLysAsnValIleAlaGlyIleArgGlyGlyGlyGlyGlyGlyGly 160
426 TCTAAGAAGTTTGTCAAGATGTGTATACAGGATTAAGGAAAGCAAGAGAGAAATCTTAC 485
161 TyrIleValGlnGlnGlyLysAsnTyrArgGlySerLeuThrHisCysArgIleArgGly 180
486 TACATCGTCGAGAGAGAGAGACACAGACAGGAAATCTTACCCATGCGAGATGCGGGT 545
181 GlyMetLeuAlaMetProLysAspGluAlaAlaAsnThrLeuIleAlaAspTyrValAla 200
546 GGAATGCTAGCCATGCCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 605
201 LysSerGlyPhePheArgValPheIleGlyValAsnAspLeuGluArgGlyGlyGlyGly 220
606 AAGAGTGGCTTCTTGGGCTGTTCATGCGCTGAATGACCTTGAAGAGGAGGAGACATAC 665
221 MetPheThrAspAsnThrProLeuGlnAsnTyrSerAsnTyrAsnGlnGlyGlyProSer 240
666 ATGTTTCACAGACACACATCTCCATGACAGACTATACCACTGGAATGAGGGGAGACCCACC 725
241 AspProTyrGlyHisGlnAspCysValGlyMetLeuSerSerGlyArgTyrAsnAspThr 260
726 GACCCCTATGTCATGAGAGACTGTGTGAGATGCTGAGCTGCGAGATGAGATGACACA 785
261 GluCysHisLeuThrMetTyrPheValCysGluPheIleLysLysLysLysLysLysLys 277
786 GAGTGCACATCTTACATGATCTTGTGTGATTCATCAAGAAAGAAAG 836

RESULT 3
AX376032 1016 bp DNA linear PAT 01-MAR-2002
LOCUS Sequence 99 from Patent W00168848.
DEFINITION AX376032
ACCESSION AX376032
VERSION AX376032.1 GI:19170408
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
Baker, R.P., Chen, J., Desnoyers, L., Goddard, A., Godowski, P.J.,
Gurney, A.L., Pan, J., Smith, V., Watanabe, C.K., Wood, W.I. and
Zhang, Z.
TITLE
Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL
Patent: WO 0168848-A 99 20-SEP-2001;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1..1016
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 312 a 197 c 261 g 246 t
ORIGIN
Alignment Scores:
Pred. No.: 1,65e-114 Length: 1016
Score: 1472.00 Matches: 275
Percent Similarity: 99.64% Conservative: 1
Best Local Similarity: 99.28% Mismatches: 1
Query Match: 99.19% Indels: 0
Gaps: 0
US-09-600-932-2 (1-277) x AX376032 (1-1016)
1 MetanaglYPhealSerLeuLeuArgArgAnGlnPheleLeuLeuValLeuPheLeu 20
22 ATGAAATGGCTTTCATCTCTGCTGGAAGAAACCAATTATCTCTCTGTAATTTCTT 81
21 LeuGlnIleGlnSerLeuGlyLeuAspIleAspSerArgProThrAlaGluValCysAla 40
82 TTGCAAAATTCAGAGCTGGCTGCTGATATGATGATGATGATGATGATGATGATGATGAT 141
41 ThrHisThrIleSerProGlyProGlyAspAspGlyGlyGlyGlyGlyGlyGlyGly 60
142 ACACACACAAATTCACCAAGACCCAAAGAGATGATGATGATGATGATGATGATGATGAT 201
61 GluGlyLysHisGlyLysValGlyArgMetGlyProGlyGlyLysGlyGlyGlyGlyGly 80
202 GAGGAAAGACATGCGAAAGTGGAGCATGGGGCCGAAAGAAATTAAGAGAGAACTGGGT 261
81 AspmetGlyAspArgGlyAsnIleGlyLysThrGlyProIleGlyLysGlyLysAspLys 100
262 GATATGGGAGATCGAGCAATATGCAAGACTGGGCCCATTTGGGAGAGAGAGGTGACAAA 321
101 GlyLysLysGlyLeuLeuGlyLysProGlyLysGlyLysGlyLysGlyLysGlyLys 120
322 GGGGAAAGAGTTTCTTGGATACCTGGAATACCTGGAAGAAAGCAAGCAAGTACTGTGTGAT 381
121 CysGlyArgTyrArgLysPheValGlyGlnLeuAspIleSerIleAlaArgLeuLysThr 140
382 TTGGAAGATACCGGAAATTTTGTGACAACTGATATGATATGATATGATATGATATGAT 441
141 SerMetLysPheValLysAsnValIleAlaGlyIleArgGlyGlyGlyGlyGlyGlyGly 160
442 TCTATGAAGTTTGTCAAGATGTATAGACAGGATTAAGGAAAGCAAGAAATCTTAC 501
161 TyrIleValGlnGlnGlyLysAsnTyrArgGlySerLeuThrHisCysArgIleArgGly 180
502 TACATCGTCGAGAGAGAGAGAACTACAGGAAATCTTACCCATGCGAGATGCGGGT 561
181 GlyMetLeuAlaMetProLysAspGluAlaAlaAsnThrLeuIleAlaAspTyrValAla 200
562 GGAATGCTAGCCATGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
201 LysSerGlyPhePheArgValPheIleGlyValAsnAspLeuGluArgGlyGlyGlyGly 220
622 AAGAGTGGCTTCTTGGGCTGTTCATGCGCTGAATGACCTTGAAGAGGAGGAGACATAC 681
221 MetPheThrAspAsnThrProLeuGlnAsnTyrSerAsnTyrAsnGlnGlyGlyProSer 240
682 ATGTCACAGACACACATCTCCATGAGAACTATACCACTGGAATGAGGGGAGACCCACAC 741

OY 241 AspProtyrGlyHisGluAspCysValGluMetLeuSerSerGlyArgTyrPasnAspThr 260
 |||||||
 Db 742 GACCCATGATGTCATGAGACTGTCGAGATGTCGAGCTGTCGACATGGAATGACACA 801
 |||||||
 OY 261 GluGlyHisLeuThrMetTyrPheValCysGluPheLeuLysLysLys 277
 |||||||
 Db 802 GAGTGCATCTTACCATGTCCTGTCGAGCTTCATCAAGAAAGAAAG 852
 |||||||
 RESULT 4
 AC080033 141262 bp DNA linear PRI 20-FEB-2002
 LOCUS Homo sapiens chromosome 8, clone RP11-885J16, complete sequence.
 DEFINITION AC080033
 AC080033.10 GI:18767530
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 1 (bases 1 to 141262)
 Bliren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 8, clone RP11-885J16
 Unpublished
 2 (bases 1 to 141262)
 Bliren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bede,F., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
 Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hago,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Lacomque,K.,
 Lamaares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacDonald,P., Margulis,N., McCarthy,M., McEwan,P., McKernan,K.,
 McPheters,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,V.,
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Sognez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 JOURNAL
 Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 141262)
 Bliren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hago,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacomque,K., Lamaares,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
 MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., Meldrim,J., Menus,L., Mihova,T.,
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 JOURNAL
 Submitted (15-FEB-2002) Whitehead Institute/MIT Center for Genome

REFERENCE
 AUTHORS
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 141262)
 Bliren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hago,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacomque,K., Lamaares,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
 MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., Meldrim,J., Menus,L., Mihova,T.,
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 JOURNAL
 Submitted (20-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 20, 2002 this sequence version replaced gi:18653568.
 All repeats were identified using RepeatMasker:
 Smt, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10933
 Center clone name: 885_J_16

 FEATURES
 SOURCE
 Location/Qualifiers
 1. 141262
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-885J16"
 /clone_1lb="RP11-11 Human Male BMC"
 31. 35
 /note="<30 qual SNGL region"
 76. 82
 /note="<30 qual SNGL region"
 88. 297
 /note="<30 qual SNGL region"
 /rpt_family="MIR"
 113. 120
 /note="<30 qual SNGL region"
 1024. 1029
 /note="<30 qual SNGL region"
 1078. 1082
 /note="<30 qual SNGL region"
 1428. 1557
 /note="<30 qual SNGL region"
 /rpt_family="MIR"
 1470. 1476
 /note="<30 qual SNGL region"
 1500. 1507
 /note="<30 qual SNGL region"
 1519. 1528
 /note="<30 qual SNGL region"
 1547. 1556
 /note="<30 qual SNGL region"
 1565. 1579
 /note="<30 qual SNGL region"
 1643. 1648
 /note="<30 qual SNGL region"


```

VERSION      AC023487.10   GI:13352736
KEYWORDS     HTGS_PHASE; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE       ORGANISM
             Homo sapiens.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 182475)
AUTHORS      Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N., Glukhov,S., Hansen,N., Hyman,R., Mo,J., Marathe,R., Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Webb,C., Wilhelmy,T., Yu,S. and Davis,R.W.
JOURNAL      2 (bases 1 to 182475)
REFERENCE    Submitted (14-FEB-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT      On Mar 16, 2001 this sequence version replaced gi:1332478.

----- Genome Center -----
Center: Stanford DNA Sequencing and Technology Development Center
Center code: SDSFDC
Web site: http://sequence-www.stanford.edu/group/human/contact: hum-info@sequence.stanford.edu
----- Project Information -----
Center project name: 844
Center clone name: RP11-164H21
----- Summary Statistics -----
Sequencing Vector: M13mp18; X02513; 98% of reads
Sequencing Vector: plasmid; plasmid.accession; 2% of reads
Chemistry: Dye-primer; 0% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180260 bases at least Q40
Consensus quality: 180441 bases at least Q30
Consensus quality: 180507 bases at least Q20
Insert size: 178614; agarose-fp
Insert size: 182275; sum-of-contigs
Quality coverage: 8.1x in Q20 bases; agarose-fp
Quality coverage: 7.9x in Q20 bases; sum-of-contigs.
* NOTE: This is a working draft sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* *
* 1 42072: contig of 42072 bp in length
* 42073 42172: gap of unknown length
* 42173 109254: contig of 67082 bp in length
* 109255 109355: gap of unknown length
* 109355 182475: contig of 73121 bp in length.
Location/Qualifiers
1. 182475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-164H21"
/clone.lib="NPC1 human BAC library 11"
1. 42072
/note="assembly_name:Contig9
clone_end:17"
42173. 109254
/note="assembly_name:Contig10"
109355. 182475
/note="assembly_name:Contig11
clone_end:SP6"

```

BASE COUNT	56447 a	33954 c	34262 g	57610 t	202 others
ORIGIN					
Alignment Scores:					
Pred. No.:	6.38e-49	Length:	182475		
Score:	720.50	Matches:	158		
Percent Similarity:	56.15%	Conservative:	11		
Best Local Similarity:	52.49%	Mismatches:	39		
Query Match:	48.55%	Indels:	93		
DB:	2	Gaps:	6		
US-09-600-932-2 (1-277) x AC023487 (1-182475)					
QY	66	LysValGlyArgMetGlyProLysGlyLeuLysGlyLeuGlyAspMetGlyAspArg	85		
DB	6734	AAACTCGGGGAAGGAGTCACTGCCATGTCATTCACAAAGGCAFTTGGAAAGACTTC-----	6684		
QY	86	GlyAsnIleGlyLysThrGlyProIleGlyLysLysGlyAspLys-----Gly	101		
DB	6683	---TTATAGGGGGAATAATATTCACACTGAGATCTAAAGGATGAGTAGAAGTTTACCAGCT	6627		
QY	102	GIuysGlyLeuLeuGlyIleProGlyLulysGlyLysAlaGlyThValCysAspCys	121		
DB	6636	AGAAAGACTTCTTAAGCATAAAGGCTTGGAAGTATAGAGAAAAGGCGCTGCTCAGT	6567		
QY	121	-----	121		
DB	6566	AGCCACAGTATAGCTAAAGCATAACTATGTTGGGAATATAGATAGATGGGCTGCG	6507		
QY	121	-----	121		
DB	6506	TGAAAGTTCATGTAACATGCTCTTGTGCACTTGTGCTGCCAGGTGTGAGAAATTGA	6447		
QY	122	-----GlyArgTyrArgLysPheValGlyIleuAspIle	133		
DB	6446	GGAGAAAAGGAAGGAGAAACAAGGATGATATGCTCTATTTGCTAACTAGAGGTG	6387		
QY	134	SerIleAlaArg-----	137		
DB	6386	GGGGTAATTAATCTGTGACTACCTCATATATAGGCGCTTTGAAAATAATTTGTGAATAT	6327		
QY	138	-----LeuLysThrSerMetLysPheVal-----	145		
DB	6326	GAATTAATAATGTAATCTGGCAATATCATATATTTTGTGCTTATTCCTTTTGAG	6267		
QY	146	---LysAsn-----ValIleAlaGlyIleArgGluThrGlu	156		
DB	6266	ATAGCTATGATACCTCTTTTCTCTCCCTGACAGATAGGAGGATTAAGGAATAACTGA	6207		
QY	157	GIuysPheTyrTyrIleValGlnGlnGluLulysAsnTyrArgLysLeuThrHisCys	176		
DB	6206	GAGAAATTCCTCTACATCTGTGCAGAAAGAACTACAGGAAATCCCTTAACCCATGCG	6147		
QY	177	ArgIleArgGlyGlyMetLeuAlaMetProLysAspGluAlaAlaAsnThrLeuIleAla	196		
DB	6146	AGGATGGGGGTGGATGCTAGCCATGCGCAAGATGAGAGTGCACACACTCATGCT	6087		
QY	197	AspTyrValAlaLysSerGlyPhePheArgValPheIleGlyValAsnAspLeuGluArg	216		
DB	6086	GACTATGTTGCCAAGAGTGGCTTCTTGGGTGTTTCATTTGGCGTAATGACCTTGAAGG	6027		
QY	217	GIuGIyGlnTyrMetPheThrAspAsnThrProLeuGlnAsnTyrSerAsnTyrAsnGlu	236		
DB	6026	GAGGACAGTCACTGTTCCACAGACACACTCCACTCGAAGACTATAGCACTGGAAATGAG	5967		
QY	237	GlyIuProSerAspProTyrGlyHisGluAspCysValGluMetLeuSerSerGlyArg	256		
DB	5966	GGGGAAACCAACGACCCCTATGGTCATGAGAGACTGTGGAGATGCTGAGCTTGCGAGA	5907		
QY	257	TrpAsnAspThrGluCysHisLeuThrMetTyrPheValCysGluPheIleLysLysLys	276		
DB	5906	TGGAATGACACAGAGGTGCATCTTAACCATGTACTTTGTCTGTGATTCATCAACAAAGAA	5847		

[illegible]

QY	157	glushysherYrTlLevalGlnGlnluysanYrAargluserLeuThrNisCys	176
Db	517	AGCAAGATCTACCGCTGGTGTGAAGAGAGAGAACGGCTACGGCGAGCCAGCTCTCTGC	576
QY	177	ArglleargglgylMetLeuAlaMetProLysAspGluAlaAlaAsnThrLeuIleAla	196
Db	577	CAGGGCGGGGGGACAGCTGATGGCCCAAGAGACAGGCGCTGCCAATGGCTGATGGCC	636
QY	197	AspYrYAlaAlaLysSerGlyPhePheArgYAlaIleIleGlyValAlaAsnAspLeuGluArg	216
Db	637	GCAATACCTGGCGCAAGCGCGGCTGGCCCGTGTCTTCATCGGCATCAACAGACTGGAAGAG	696
QY	217	GlulGlyLglnYrMetPheThrAspAsnThrProLeuGlnAsnYrSerAsnThrAspGlu	236
Db	697	GAGGGCGCCCTTCGCTACCTGACCACTCCGCCATCGGGACCTTCAACAGTGGCGACAC	756
QY	237	GlycLubProSerAspProYrGlyLysGlnAspCysValGluMetLeuSerSergYlArg	256
Db	757	GGTGAAGCCCAACATAGCTACGACGACGAGAGAGACATCGCTGAGATGATGGCTCGGGGCGC	816
QY	257	TrpAsnAspThrGlyCysHisLeuThrMetYrPheValCysGluPheIleYlGlyLys	275
Db	817	TGGAAGACAGTGGCGCTCCACACCAACCATGTACTTCATGTTGATTGACAGAGAG	873
RESULT 7			
AX454582			
LOCUS	AX454582	1238 bp	DNA linear PAT 06-JUL-2002
DEFINITION	Sequence 167 from Patent WO0208284.		
ACCESSION	AX454582		
VERSION	AX454582.1	GI:21713915	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A., Godowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J., Paoni,N.F., Stephan,J.P., Wetanabe,C.K., Williams,P.M., Wood,W.I. and Ye,W.		
TITLE	Compositions and methods for the diagnosis and treatment of disorders involving angiogenesis		
JOURNAL	Patent: WO 0208284-A 167 31-JAN-2002; Genentech, Inc. (US); Baker, Kevin P. (US); Ferrara, Napoleone (US); Gerber, Hanspeter (US); Gerritsen, Mary E. (US); Goddard, Audrey (US); Godowski, Paul J. (US); Gurney, Austin L. (US); Hillan, Kenneth J. (US); Marsters, Scott A. (US); Pan, James (US); Paoni, Nicholas F. (US); Stephan, Jean-Philippe F. (US); Wetanabe, Colin K. (US); Williams, P. Mickey (US); Wood, William I. (US)		
FEATURES			
source	Location/Qualifiers		
	1..1238		
BASE COUNT	293 a 321 c 389 g 235 t		
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606"		
Alignment Scores:			
Score: NO:	2.13e-51	Length:	1238
Percent Similarity:	719.50	Matches:	130
Best local Similarity:	67.74%	Conservative:	59
Query Match:	46.59%	Mismatches:	81
	48.48%	Indels:	9
DB:	6	Gaps:	3
US-09-600-932-2 (1-277) x AX454582 (1-1238)			
QY	3	glyPheAlaSerLeuArgArgAsnGlnPheIleLeuValLeuPheLeuGln	22
Db	46	CGTCTCCTCGCTGGCGCTCAGAGATAGAGGGGGAATGTGGCCGTGGGCGCTTCTAATACG	105
QY	23	IleGlnSerLeuGlyLeuAspIleAspSerArgProThrAla-----GluValCys	39

Douthwaite, R.J., Draper, H., Dugan-Rocha, S., Durbin, R.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P.,
 Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Corbell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kravtsov, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, Z., Lichte, O., Lie, C., Liu, J., Liu, R., Liu, R., Ma, J.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Mageshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mathney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, A., Nguyen, N.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okunolu, G.,
 Otagunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Qullies, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N., Thomas, S.,
 Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 267118)
 Worley, K.C.
 Direct Submission
 Submitted (06-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 267118)
 Worley, K.C.
 Direct Submission
 Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 24, 2002 this sequence version replaced g1:17063112.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center Project name: TUDS
 Center Clone name: CH230-2F22
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap, version 0.990329
 Consensus quality: 216265 bases at least Q40
 Consensus quality: 216981 bases at least Q30
 Consensus quality: 217528 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 46 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1066: contig of 1066 bp in length
 * 1067 1166: gap of unknown length
 * 2355: contig of 1189 bp in length
 * 2356 2455: gap of unknown length
 * 2456 3677: contig of 1222 bp in length
 * 3678 3777: gap of unknown length
 * 3778 5145: contig of 1368 bp in length
 * 5146 5245: gap of unknown length
 * 5246 6443: contig of 1198 bp in length
 * 6444 7700: gap of unknown length
 * 7701 7800: contig of 1157 bp in length
 * 7801 9178: gap of unknown length
 * 9179 9278: contig of 1378 bp in length
 * 9279 10397: gap of unknown length
 * 10398 10497: contig of 1119 bp in length
 * 10498 11814: gap of unknown length
 * 11815 11914: contig of 1317 bp in length
 * 11915 12984: gap of unknown length
 * 12985 13084: contig of 1070 bp in length
 * 13085 14314: gap of unknown length
 * 14315 15533: contig of 1230 bp in length
 * 15534 15634: gap of unknown length
 * 15635 16692: contig of 1119 bp in length
 * 16693 16792: gap of unknown length
 * 16793 17990: contig of 1059 bp in length
 * 17991 18148: gap of unknown length
 * 18149 19248: contig of 1198 bp in length
 * 19249 20311: gap of unknown length
 * 20312 20411: contig of 1063 bp in length
 * 20412 21611: gap of unknown length
 * 21612 21711: contig of 1200 bp in length
 * 21712 22800: gap of unknown length
 * 22801 22900: contig of 1089 bp in length
 * 22901 23960: gap of unknown length
 * 23961 24060: contig of 1060 bp in length
 * 24061 25114: gap of unknown length
 * 25115 25214: contig of 1054 bp in length
 * 25215 26371: gap of unknown length
 * 26372 26471: contig of 1157 bp in length
 * 26472 27623: gap of unknown length
 * 27624 27723: contig of 1151 bp in length
 * 27724 28851: gap of unknown length
 * 28852 28951: contig of 1129 bp in length
 * 28952 30133: gap of unknown length
 * 30134 31500: contig of 1182 bp in length
 * 31501 31600: gap of unknown length
 * 31601 32791: contig of 1267 bp in length
 * 32792 32891: gap of unknown length
 * 32892 34072: contig of 1190 bp in length
 * 34073 34172: gap of unknown length
 * 34173 35173: contig of 1182 bp in length
 * 35174 35273: gap of unknown length
 * 35274 36350: contig of 1001 bp in length
 * 36351 36450: gap of unknown length
 * 36451 38017: contig of 1077 bp in length
 * 38018 39244: gap of unknown length
 * 39245 39344: contig of 1567 bp in length
 * 39345 40695: gap of unknown length
 * 40696 41966: contig of 1127 bp in length
 * 41967 42066: gap of unknown length
 * 42067 43627: contig of 1127 bp in length
 * 43628 45091: gap of unknown length
 * 45092 45191: contig of 1352 bp in length
 * 45192 46386: gap of unknown length
 * 46387 50256: contig of 1461 bp in length
 * 50257 50356: gap of unknown length
 * 50357 59360: contig of 3770 bp in length
 * 59361 59360: gap of unknown length
 * 59361 59360: contig of 9004 bp in length

chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis Mo. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPI1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Errenge, E., Tatenoe, M., Catanesse, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RPI1-178E6, 2000 bp overlap. Actual start of this clone is at base position 1 of RPI1-568H24.

The sequence between 66093 to 66578 and 104506 to 104590 is covered only by PCR products from clone DNA. The sequence contains a dinucleotide (TC) run from 65513 to 65634 in which the exact length is unknown. The sequence contains a dinucleotide (TC) run from 104386 bp to 104631 bp in which the exact length is unknown. The sequence from base position 4458 to 6187 can not be guaranteed due to a tandem repeat.

FEATURES

source

```

1. 139357
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="2"
   /map="2"
   /clone="RPI1-568H24"
   /clone_id="RPI1-11"
   /rpt_family="AT-rich"
   383..409
   /rpt_family="Achoho"
   876..950
   /rpt_family="Achoho"
   951..1248
   /rpt_family="Alu"
   1249..1339
   /rpt_family="Achoho"
   1340..1514
   /rpt_family="MERL-type"
   1611..1965
   /rpt_family="Achoho"
   1973..2279
   /rpt_family="Alu"
   2324..2843
   /rpt_family="L1"
   2844..3050
   /rpt_family="MERL-type"
   2943..2958
   /note="similar to Homo sapiens EST A1597790 (NID:94606838)
   tr92h02.x1"
   3051..3174
   /rpt_family="L1"
   3175..3483
   /rpt_family="Alu"
   3182..3206
   /note="similar to Homo sapiens EST BG178637
   (NID:912685340)"
   3484..4227
   /rpt_family="L1"
   4235..4330
   /rpt_family="(T)n"

```

```

repeat_region 4312..4987
   /rpt_family="SVA"
misc_feature 4571..6524
   /note="cpg_island (%GC=74.6, c/e=0.80, #cpgs=159)"
repeat_region 4978..5152
   /rpt_family="(CCGC)n"
repeat_region 5163..5560
   /rpt_family="SVA"
repeat_region 5249..5273
   /rpt_family="GC-rich"
repeat_region 5593..6486
   /rpt_family="SVA"
repeat_region 6443..6529
   /rpt_family="GA-rich"
repeat_region 6531..6732
   /rpt_family="L1"
repeat_region 6733..7033
   /rpt_family="Alu"
repeat_region 7034..7249
   /rpt_family="L1"
repeat_region 7250..7554
   /rpt_family="Alu"
repeat_region 7555..8490
   /rpt_family="L1"
repeat_region 8528..8726
   /rpt_family="Achoho"
misc_feature 8666..8669
   /note="match to EST BF314952 (NID:911263123)"
misc_feature 9031..9360
   /note="match to EST BF314952 (NID:911263123)"
misc_feature 9047..9360
   /note="match to EST BE259985 (NID:99130920)"
misc_feature 9068..9069
   /note="match to EST BF316717 (NID:911265074)"
misc_feature 9068..9069
   /note="match to EST BF316717 (NID:911265074)"
misc_feature 9144..9147
   /note="match to EST BF3198782 (NID:914653803)"
misc_feature 9239..9360
   /note="similar to Homo sapiens EST BF311850
   (NID:911259616)"
misc_feature 9239..9360
   /note="match to EST BF3198831 (NID:914653852)"
misc_feature 9239..9360
   /note="similar to Homo sapiens EST BF3198308
   (NID:914653329)"
misc_feature 9239..9360
   /note="similar to Homo sapiens EST BF3198502
   (NID:914653523)"
misc_feature 9240..9360
   /note="match to EST BF206235 (NID:911099821)"
misc_feature 9240..9360
   /note="match to EST BF311988 (NID:911259763)"
misc_feature 9240..9360
   /note="match to EST BF319896 (NID:914654917)"
misc_feature 9240..9360
   /note="similar to Homo sapiens EST BF313026
   (NID:911260952)"
misc_feature 9313..9360
   /note="match to EST BF311981 (NID:911259754)"
misc_feature 9313..9360
   /note="similar to Homo sapiens EST BF316085
   (NID:911264380)"
misc_feature 9313..9360
   /note="similar to Homo sapiens EST BF3198293
   (NID:914653314)"
misc_feature 9322..9360
   /note="match to EST BF316717 (NID:911265074)"
misc_feature 9322..9360
   /note="match to EST BF3198782 (NID:914653803)"
misc_feature 9322..9360
   /note="similar to Homo sapiens EST BF311850
   (NID:911259616)"
misc_feature 9333..9360
   /note="match to EST BE313758 (NID:99134050)"

```


Db	163	AGGAGGGGCAATGCTTCCCTCCCTCCCTCCATATTTGGTCCGTGCACACAGATCCCAAGT	222
Qy	26	----IeuclIyeuAspIleAsp-----	31
Db	223	TTTCCTGGCCGGAAGAAATGGATGTATTCTGAGAAAACCTCACCGATCCCTGCACCTG	282
Qy	32	-----SerArgProThrAlaGluValCysAlaThrHisThrIleSerProGlyProlys	49
Db	283	GTTCGTGTGTGCTCCACACAGAGATAGCTTCCTGGTGTATGATGGACGAGATGGGAAAAGA	342
Qy	50	GLYAspAspGlyIleuLysGlyAspProGlyIleuGluGlyLysHisGlyLysValIleArg	69
Db	343	GGCCCCCAGAGGAGAAAGGGGGATCCAGGGCCTCAGGTATGCTGTAGCCTGGACCACTGGAGA	402
Qy	70	MetGlyProlys-----	73
Db	403	GAGGGCCCCCTCAGGGAGGAGGAGGAGCATGGGACCTCCAGGCACACAGGCCCCAAAGGA	462
Qy	73	-----	73
Db	463	GAACCTGGGCCCCGAAGGAGAGAGTGGGTGCCCCACGAGATGCCGGGTCCCGACGCCCGCA	522
Qy	74	GLYIleLysGlyIleuLysGlyAspMetClyAspArgGlyAsnIleGlyLysThrGlyPro	93
Db	523	GGTCTCAAAGGAGAGAGAGATACCCCTGGGGCTGTGGTGAAGCATCGGTCCACAGGGACT	582
Qy	94	IleGlyLysLysGlyAspLysGlyIleuLysGlyLysGlyIleProGlyIleuLysGly	113
Db	583	TCAGTGGCATGGGTCCGCCACGAGCATGAAGGAGACAGAGAGTGTATCTCCGGGAAAAAGGA	642
Qy	114	LysAlaGly-----ThrValCysAspCysGlyArgTrpArgLysPheValIleGlyLeu	131
Db	643	GCAAGGGGGGAGACTTCACTCTTGAAGGTGTATCTCTTAAGCGAGCGGATGAGAACTTA	702
Qy	132	AspIleSerIleAlaArgLeuLysThrSerMetLysPheValLysAsnValIleAlaGly	151
Db	703	GAGGAGAAAGTTCACACGCTC-----CAAAACATAGTCACTAG	741
Qy	152	IleArgGlu-----ThrGluLysPheTrpTrpIle	162
Db	742	TACAGGAAGCGGTCTTCTCCGTGATGGCCAGGCTGCGGGAGAAAGATCTTCAAGACG	801
Qy	163	ValGluGluGluLysAsnTrpArgGluSerIleuThrHisCysArgIleArgGlyLysMet	182
Db	802	GCAGGTGCTGTAAGATCATATTGCGATCAAGACAGACTCTGCAGAGAGGCTTAAGGACAG	861
Qy	183	LeuAlaMetProLysAspGluAlaAlaAsnThrIleuIleAlaAspTrpVal--AlaLys	201
Db	862	CTGGCTTCCTCCACGCTCTTCAGCCGAGAGACAGGCCCGACACACAGCTGTACAGCCCAAG	921
Qy	202	SerGlyPhePheArgValPheIleGlyValAsnAspLeuGluArgGlyLysGlyLysMet	221
Db	922	AACAAAG-----CATGCTTACCTAGCATGATATACATCTCCAAAGAGGCAAGTTCACC	975
Qy	222	PheThrAspAsnThrProLysGluAsnTrpSerAsnTrpAsnGluGlyLysProSerArg	241
Db	976	TACCCAAAGGGGGGGTCACTG---GACTATTCCAACTGCGCCCGGGGAGGCCCAACAT	1033
Qy	242	-----ProTyrGlyHisGluAspCysValGluMetLeuSerSerGlyArgTrpAsn	258
Db	1033	AGGGCAAAAGACGAGAGGCCAGAGAACTGTTGTGAGATCTTTCGGATGCGAATGGAT	1092
Qy	259	AspThrGlyCysHisLysLeuThrMetLysPheValGlyGluPhe	272
Db	1093	GACATAGATGACAGGAGAGAGCGCTGTGTGTGTGTGTG 1134	
RESULT 14			
BLSPD	1454 bp	mrna	linear
LOCUS	B.taurus	mrna	linear
DEFINITION	B.taurus mRNA for lung surfactant protein D.		
ACCESSION	X75911		
VERSION	X75911.1		
KEYWORDS	lung surfactant protein D.		

THIS PAGE BLANK (USPTO)